

**Белорусский государственный университет  
Факультет социокультурных коммуникаций**

**Кафедра компьютерной лингвистики и лингводидактики**

**Реферат дипломной работы  
«Автоматизация создания и пополнения параллельного корпуса  
текстов»**

**Козловская Анна Леонидовна,  
Руководитель Воронович Валерий Викторович**

**2015 год**

## РЕФЕРАТ

**Дипломная работа:** страниц – 90, в том числе иллюстраций – 12, библиографический список – 30 источников, приложений - 6.

**Ключевые слова:** корпусная лингвистика, лингвистический корпус текстов, параллельный корпус текстов, программный продукт.

**Объектом** исследования является параллельный корпус текстов.

**Предметом** исследования являются особенности и этапы создания англо-русского параллельного корпуса текстов.

**Цель** работы – разработать программное обеспечение для автоматического создания параллельного корпуса текстов.

При выполнении работы использованы **методы** статистического и сравнительного анализа.

В процессе работы проведены следующие **исследования и разработки:** программа OTSplit, параллельный корпус текстов, созданный при помощи данной программы на основе романа «Джейн Эйр».

Элементом **научной новизны** полученных результатов является алгоритм автоматического создания параллельного корпуса текстов.

Область возможного **практического применения** – при обучении иностранному языку и переводу в школах и университетах; в лабораториях, работающих в области компьютерной лингвистики и др.

Результатом внедрения явился готовый англо-русский параллельный корпус текстов на 40 тысяч слов по роману «Джейн Эйр».

**Социально-экономическая значимость:** программа значительно сокращает временные затраты на создание параллельных корпусов текстов.

В ходе выполнения дипломной работы результаты прошли **апробацию** на 72-ой научной конференции студентов и аспирантов Белорусского государственного университета в виде доклада и презентации. Автор работы подтверждает, что приведенный в ней расчетно-аналитический материал правильно и объективно отражает состояние исследуемого процесса, а все заимствованные из литературных и других источников теоретические, методологические и методические положения и концепции сопровождаются ссылками на их авторов.

## ABSTRACT

**Diploma work:** pages – 90, including illustrations – 12, references – 30, supplement – 6.

**Key words:** corpus linguistics, linguistic corpus, parallel corpus, software program.

**Object of research:** parallel corpus.

**Subject of research:** peculiarities and stages of creation English-Russian parallel corpus.

**Aim of work:** to develop a software program for automatic creation of parallel corpus.

**Research methods:** statistical and comparative analyses.

During the research work **were created** software program OTSplit and English-Russian parallel corpus based on the novel «Jane Eyre» with the usage of this software program.

Elements of **scientific novelty** of the results is the ability to computerize a process of parallel corpus creation.

The area of possible **practical application** – teaching foreign languages and translation in schools and universities; in laboratories working in the field of computational linguistics, etc.

**The result of introduction** is the ready English-Russian parallel corpus on 40 thousand words based on the novel «Jane Eyre».

**Socio-economic importance** – the program significantly reduces the time required to create a parallel corpus.

The area of **practical approbation:** 72nd Scientific Conference of students and graduate students of the Belarusian State University.

The author confirms that all the material included in the work correctly and objectively reflects the state of the process studying, and all borrowed theoretical, methodological and methodical positions and concepts from literature and other sources are accompanied by the references to their authors.

## РЭФЕРАТ

**Дыпломная праца:** старонак – 90, у тым ліку ілюстрацый – 12, бібліяграфічны спіс – 30 крыніц, дадаткаў – 6.

**Ключавыя словы:** корпусная лінгвістыка, лінгвістычны корпус тэкстаў, паралельны корпус тэкстаў, праграмны прадукт.

**Аб'ектам** даследавання з'яўляецца паралельны корпус тэкстаў.

**Прадметам** даследавання з'яўляюцца асаблівасці і этапы стварэння англа-рускага паралельнага корпуса тэкстаў.

**Мэта** працы – распрацаваць праграмнае забеспячэнне для аўтаматычнага стварэння паралельнага корпуса тэкстаў.

Пры выкананні працы выкарыстоўваліся **метады** статыстычнага і параўнальнага аналізу.

У працэсе работы былі праведзены наступныя **даследаванні і распрацоўкі:** праграма OTSplit, паралельны корпус тэкстаў, створаны пры дапамозе дадзенай праграмы на аснове рамана «Джэйн Эйр».

Элементарам **навуковай навізны** атрыманых вынікаў з'яўляецца алгарытм аўтаматычнага стварэння паралельнага корпуса тэкстаў.

Галіна магчымага **практычнага прымянення** – пры навучанні замежнай мове і перакладу ў школах і ўніверсітэтах; у лабараторыях, якія працуюць у галіне камп'ютарнай лінгвістыкі і інш.

**Вынікам ўкаранення** з'явіўся гатовы англа-рускі паралельны корпус тэкстаў на 40 тысяч слоў па рамана «Джэйн Эйр».

**Сацыяльна-эканамічная значнасць** – праграма значна скарачае часавыя затраты на стварэнне паралельных карпусоў тэкстаў.

У ходзе выканання дыпломнай працы вынікі прайшлі **апрабацыю** на 72 навуковай канферэнцыі студэнтаў і аспірантаў Беларускага дзяржаўнага ўніверсітэта ў выглядзе дакладу і прэзентацыі.

Аўтар працы пацвярджае, што прыведзены ў ёй разлікова-аналітычны матэрыял правільна і аб'ектыўна адлюстроўвае стан доследнага працэсу, а ўсе запазычаныя з літаратурных і іншых крыніц тэарэтычныя, метадалагічныя і метадычныя палажэнні і канцэпцыі суправаджаюцца спасылкамі на іх аўтараў.